

基于随机访问的多视点视频编码模式评价模型*

杨 铀^{1,2,3}, 蒋刚毅^{1,2,3+}, 郁 梅^{1,2,4}, 曹德侠⁵

¹(中国科学院 计算技术研究所,北京 100190)

²(中国科学院 研究生院,北京 100049)

³(宁波大学 信息科学与工程学院,浙江 宁波 315211)

⁴(北京大学 视听信息处理国家重点实验室,北京 100871)

⁵(中国矿业大学 理学院,江苏 徐州 221008)

A Random Access Based Evaluation Model for Multiview Video Coding Schemes

YANG You^{1,2,3}, JIANG Gang-Yi^{1,2,3+}, YU Mei^{1,2,4}, CAO De-Xia⁵

¹(Institute of Computing Technology, The Chinese Academy of Sciences, Beijing 100190, China)

²(Graduate University, The Chinese Academy of Sciences, Beijing 100049, China)

³(Faculty of Information Science and Engineering, Ningbo University, Ningbo 315211, China)

⁴(National Key Laboratory of Machine Perception, Peking University, Beijing 100871, China)

⁵(College of Sciences, China University of Mining and Technology, Xuzhou 221008, China)

+ Corresponding author: E-mail: jianggangyi@126.com

Yang Y, Jiang GY, Yu M, Cao DX. A random access based evaluation model for multiview video coding schemes. *Journal of Software*, 2008,19(9):2313–2321. <http://www.jos.org.cn/1000-9825/19/2313.htm>

Abstract: In this paper, evaluation function of MVC (multiview video coding) random accessibility for decoder and server is first proposed based on the methods from random graph theory and hyper-space. Furthermore, a non-linear multi-purpose programming model for MVC performance evaluation is also proposed according to the constraints from practical network bandwidth, restrictions of interaction and other problems. Based on this model, optimal coding strategies for different MVC schemes under variant performance constraints are discussed. It is found that random accessibility of MVC scheme can be improved via this non-linear multi-purpose programming model.

Key words: multiview video coding; mathematical model; random graph theory; interactive service; random access

摘 要: 运用随机图论、超空间等方法,首先提出了针对编码预测结构的随机访问性能评价函数,对解码端和服务器的随机访问支持代价进行评价.进而根据实际应用中的带宽、交互限制等多种目标建立了编码策略评价的多目

* Supported by the National Natural Science Foundation of China under Grant No.60472100 (国家自然科学基金); the NSFC/KOSEF Joint Research Project under Grant No.60472100 (中韩国家基金合作研究项目); the Key Project of the Ministry of Education of China under Grant No.206059 (国家教育部科学技术研究重点项目); the Program for New Century Excellent Talents in University of China under Grant No.NCET-06-0537 (新世纪优秀人才支持计划)

Received 2006-04-21; Accepted 2007-05-23

标非线性数学模型,对编码策略在多种性能参数约束条件下如何实现最优策略进行了讨论.利用该非线性多目标规划模型对多视点视频编码策略进行了优化设计,可以获得较好的随机访问性能.

关键词: 多视点视频编码;数学模型;随机图论;交互式服务;随机访问

中图法分类号: TP391 文献标识码: A

具有交互性的多媒体技术将使人们能够在许多方面由被动观赏转化为主动欣赏.多视点视频技术作为交互式多媒体技术中的一个重要的研究领域,将在虚拟现实系统、数字电视、远程教育、远程工业控制、三维视频会议系统等诸多方面得到直接的应用^[1].作为一般单视点视频的拓展,多视点视频的形成是通过多台摄像机同时多角度的拍摄而产生的.与传统的单视点视频相比,多视点视频需要处理大量的视频数据,在运动估计和视差估计算法上也比单视点视频要复杂.多视点视频由于其应用前景广泛而引起了研究人员的关注.MPEG (moving picture expert group)和 JVT(joint video team)正联手对三维音、视频中的多视点视频进行标准化制定工作.作为用户可控制能力和可参与性的一项重要指标,MPEG 把允许用户在多视点视频的时间和视点轴上任意切换访问作为一项基本研究内容和要求放在对多视点进行研究的提案中^[2].对于从时间轴上支持用户的随机访问,多视点视频解码器必须能够在最短的时间内以最小的解码代价满足用户在单独一个视点内观看任意时刻的一帧,这与单视点视频的随机访问相同;而对于从视点轴上支持用户的随机访问,则是以最小的解码代价满足用户观看任意一个视点的视频.对于用户而言,看似简单的这两个要求在研究上却困难重重:首先,由于多视点视频的数据量庞大,多视点视频编码(multiview video coding,简称 MVC)往往是以提高压缩比率为目标而进行的.各种压缩方法的提出使得压缩性能得到了一定的满足,但编码端的高压缩比往往意味着帧间依赖关系密切,从而导致解码端随机访问性能会明显降低^[3,4];其次,研究人员对如何量化地衡量随机访问性能没有形成统一意见.一些以研究系统性能为目标的多视点视频播放实验系统提出了各自的衡量方法,文献[5]提出,在基于光线空间系统以帧的平均处理时间来衡量,文献[6]则以服务器端的平均处理器荷载来衡量,等等.这些方法从直观上能对随机访问性能进行一定程度的评价,但是所得到的参数对于不同的多视点视频序列和测试环境会发生不可预料的变化.因此,仅以这些参数描述随机访问性能是不可行的.文献[7]提出了利用最差随机访问性能来评价 GOP(group of pictures)整体随机访问性能的方法,该方法是一种无差别的评价方法,但其评价参数仅衡量 GOP 的最差随机访问性能,不利于对 MVC 预测结构中的帧间关系进行整体、全面的评价.

考虑到随机访问性能的描述参数应符合 MPEG 的要求,即以解码帧数为参数进行衡量,这样,对于不同的环境和视频序列,其差别仅为线性的常数倍数关系,从而达到无类别普适性的评价.鉴于此,本文提出了随机访问评价函数,对解码端和服务端为支持用户的随机访问所需要付出的代价进行评价,进而建立了对多视点视频编码策略进行评价的多目标非线性数学模型.该模型能够对多视点视频编码方案的随机访问性能进行先验性评价,从而允许编码端根据不同的应用环境决定采用何种视频编码方案.

1 单帧随机访问与荷载性能评价函数

设 n 表示视点个数, m 表示单一视点内的帧数, m_1 表示一个 GOP 内单一视点内的帧数.

定义 1^[8]. 图 $G=(V,E)$ 为一个二元组,其中, V 为点集, E 为边集.一条边 $e_{ij} \in E$ 当且仅当顶点 v_i 和 v_j 之间有边相连,其中, $v_i, v_j \in V$.点集 V 的势被称为图 G 的阶数,边集 E 的势被称为图 G 的大小.所有与顶点 $v_i \in V$ 相连的顶点构成 v_i 的邻域,并记做 $N(v_i)$,同时, $N(v_i)$ 的势称为 v_i 的度.将所有顶点度数相等的图称为正则图.对一个 n 阶图而言,如果任意顶点的度都是 $n-1$,则该图为 n 阶完全图.对边上赋予方向的图称为有向图.对边上赋予权重的图称为带权图.

定义 2^[9]. 随机图 $G=(V,E,\Omega)$ 为一个三元组,其中,以 V 为点集、 E 为边集的图 G ,其边集定义在概率空间 Ω 中,即 $P(e_{ij} \in E) = p_{ij} (0 < p_{ij} < 1, \sum p_{ij} = 1)$.当 G 是一个带权有向图时,则 G 是一个随机的带权有向图.用 P 表示 G 中的路.

定义 3^[9]. 用有限顶点集 $V = \{v_i | 1 \leq i \leq n \times m\}$ 表示多视点视频序列中将要被处理的帧,则超空间 $H=(V,E)$,其中,

$E=(E_i), i \in I, E_i$ 为具有视差-运动估计关系的顶点子集, I 是 H 的指标集.

用户在由多视点视频序列所构成的超空间中接受视频服务,在未知视频内容的情况下,其对超空间中任意一帧的访问概率是相等的.令 X_f 为单帧的访问代价随机变量,以 x_i 表示在解码端为了显示第 i 帧所需要预先解码的帧数,以 p_i 表示用户访问第 i 帧的概率,则得到用户访问超空间中任意一帧的平均解码代价为^[8]

$$E(X_f) = \sum_{i=1}^{n \times m_1} x_i p_i \quad (1)$$

式中, $p_i = 1/n \times m_1$, x_i 的值由多视点视频编码时所采用的编码策略所决定.利用公式(1)可以对多视点编码策略在解码端支持用户随机访问单帧的随机访问性能进行评价.此外,若以 X_m 表示支持单帧访问的内存消耗随机变量,以 x_i 表示服务器端为了支持随机访问而需要开辟内存空间时,用 $E(X_m)$ 表示服务器端支持用户访问超空间中任意一帧所需要开辟的平均内存空间大小,则 $E(X_m)$ 也可以采用公式(1)进行计算.

2 交互模式下随机访问与荷载性能评价函数

2.1 复杂交互模式的定义

用户在终端显示设备上接收并观看多视点视频,会在一定程度上发生与视频服务器之间的交互,如改变视点、快进快退操作等.文献[6]定义了3种交互式操作,即任意视点切换、固定视点切换、连续视点切换.从多视点视频显示系统实现的角度,这3种交互模式易于实现,而用户却会由于交互模式过于简单而受到诸多限制.为了适应复杂的用户交互情况,本文在此基础上进一步定义了3种复杂的交互模式.

定义 4. 用 $V_i T_j$ 表示第 i 视点第 j 时刻的帧,其中, $1 \leq i \leq n$ 且 $1 \leq j \leq m_1$, 定义下列3种复杂交互模式:

- (1) 相邻视点切换模式:在以正常速度播放仍在继续的同时,允许用户从当前视点切换到左或者右相邻的视点进行观看,在连续切换过程中允许非单一方向,即如果当前帧为 $V_i T_j$, 则下一帧为 $V_{i \pm 1} T_{j+1}$.
- (2) 任意视点切换模式:用户在正常的观看模式或者暂停模式下,可以在各个视点之间(允许非相邻视点)进行任意切换,即如果当前帧为 $V_i T_j$, 则下一帧为 $V_{i+k} T_{j+1}$, 其中, $1 \leq i \pm k \leq n$.
- (3) 空间任意切换模式:允许用户在整个多视点视频超空间中从当前帧跳转到任意指定的帧进行观看,跳转过程不论视点或时间距离.

相邻视点切换模式和任意视点切换模式如图1所示.对于空间任意切换模式,即在允许用户进行(1)和(2)两种交互模式的同时,还允许用户进行任意的快进、快退操作,并且在快进快退的同时,也允许进行视点切换.若以图 G 的形式表示,则该模式对应的图为 $n \times m$ 阶完全图 $K_{n \times m}$.从图论分析的角度,无向图即为每对顶点之间同时连接着正向与反向边,因此,可以在有限的超空间中把无向图拆分成正向有向图和反向有向图.从图论分析的角度,在用户进行快进或者正常观看模式时,则可以采用正向有向图进行分析;反之,在用户进行快退或者逆向观看模式时,则可以采用反向有向图进行分析.而无论正向或者反向有向图,由备选帧所构成的超空间在图论上都是等价的.因此,只需要分析一种方向的有向图即可.所以,为了方便起见,本文只分析正向有向图的情况.从上面的3种交互模式的定义中可以看出,空间任意切换模式包含了任意视点切换模式,而任意视点切换模式又包含了相邻视点切换模式.空间任意切换模式包含了用户交互模式的任意情况,是最复杂的用户与服务器之间的交互模式.然而,无论是哪种交互模式,其分析方法在图论意义下都是相同的.

在由多视点视频的一个 GOP 构成的超空间中,用户在观看的过程中对帧的访问属于概率空间 Ω 中的一个随机事件.为了方便起见,本文在 GOP 开始之前和结束之后分别加入了一个虚拟的起点和终点,如图1所示的 Start Point 和 End Point.于是,用户从起点开始至终点所观看的这些帧都可以构成有向随机图 G 当中的一条随机有向路,对 GOP 的随机访问性能评价将建立在 G 的基础上.MPEG 工作文档中指出,性能优劣应以衡量解码帧的数量为宜^[2].采用衡量解码帧的数量可以避开运行测试环境和条件的不同而带来的差异,在不同环境下的测量结果最多只会相差一个常数倍,因此能够体现无类别的评价,所以是可靠的.

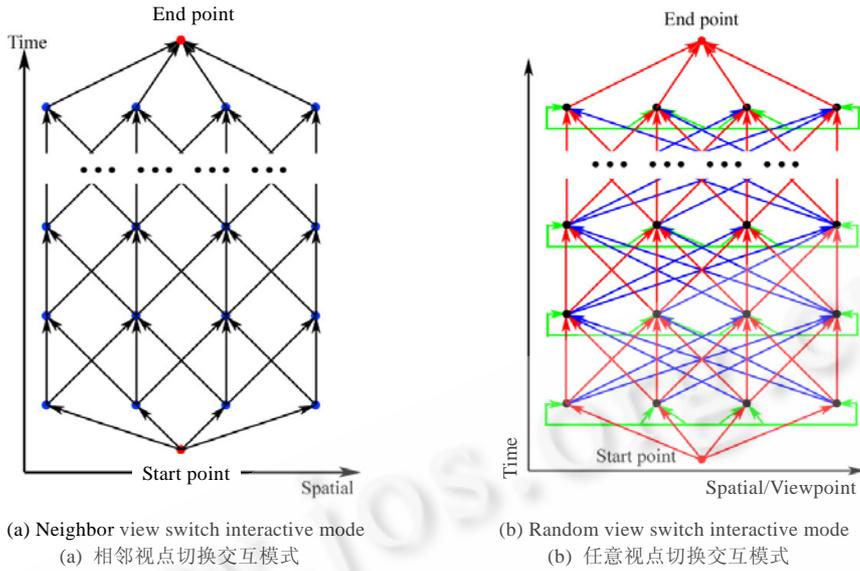


Fig.1 Two interactive modes between multiview video server and client

图 1 两种常用的多视点视频服务器和用户之间的交互模式

2.2 复杂交互模式下的随机访问性能评价函数

在未知视频内容的情况下,用户在观看由 n 个视点构成的多视点视频节目时,其选择任意一个视点进行观看的概率是等概率事件,即 $P(\text{选择一个视点})=1/n$.在正常观看模式条件下,即相邻视点切换和任意视点切换的交互模式下,用户保持在单一视点内进行观看的概率应比切换到相邻视点进行观看的概率要大,切换到更远视点进行观看的概率比切换到相邻视点的概率要小.在统计意义下,不失一般性,可以假设用户在观看的过程中进行视点切换的概率服从 $N(0,1)$ 分布.而对于不服从 $N(0,1)$ 分布的随机变量 $X \sim N(\mu, \sigma^2)$,可以通过式(2)进行转换:

$$\frac{X - \mu}{\sigma} \sim N(0,1) \tag{2}$$

在图 2 所示的 4 种情况中,每个图的每条边上均赋有概率值,表示从箭尾的当前帧切换到箭头所指的下一帧的概率.对于从起点开始用户访问 8 个视点的第一时刻的帧属于等概率事件,因此访问概率为 $1/8$.除此之外,本文设定保持单视点观看的概率为 0.6,切换到其他视点的概率为 0.2.对于不同的应用实例,这个切换的概率是不同的,具体依据实际需要而定,或者通过大量的用户测试进行统计确定.从当前帧选择如何切换是一个概率空间中的随机事件,而从当前帧引出的边构成了概率空间中的所有事件,所以其概率之和应为 1,这是符合概率空间要求的.进一步地,由起点至终点所构成的随机有向路也构成了用户进行切换选择观看多视点视频这一事件的概率空间,而这个事件的概率空间为 1,是进行理论研究的基本要求.

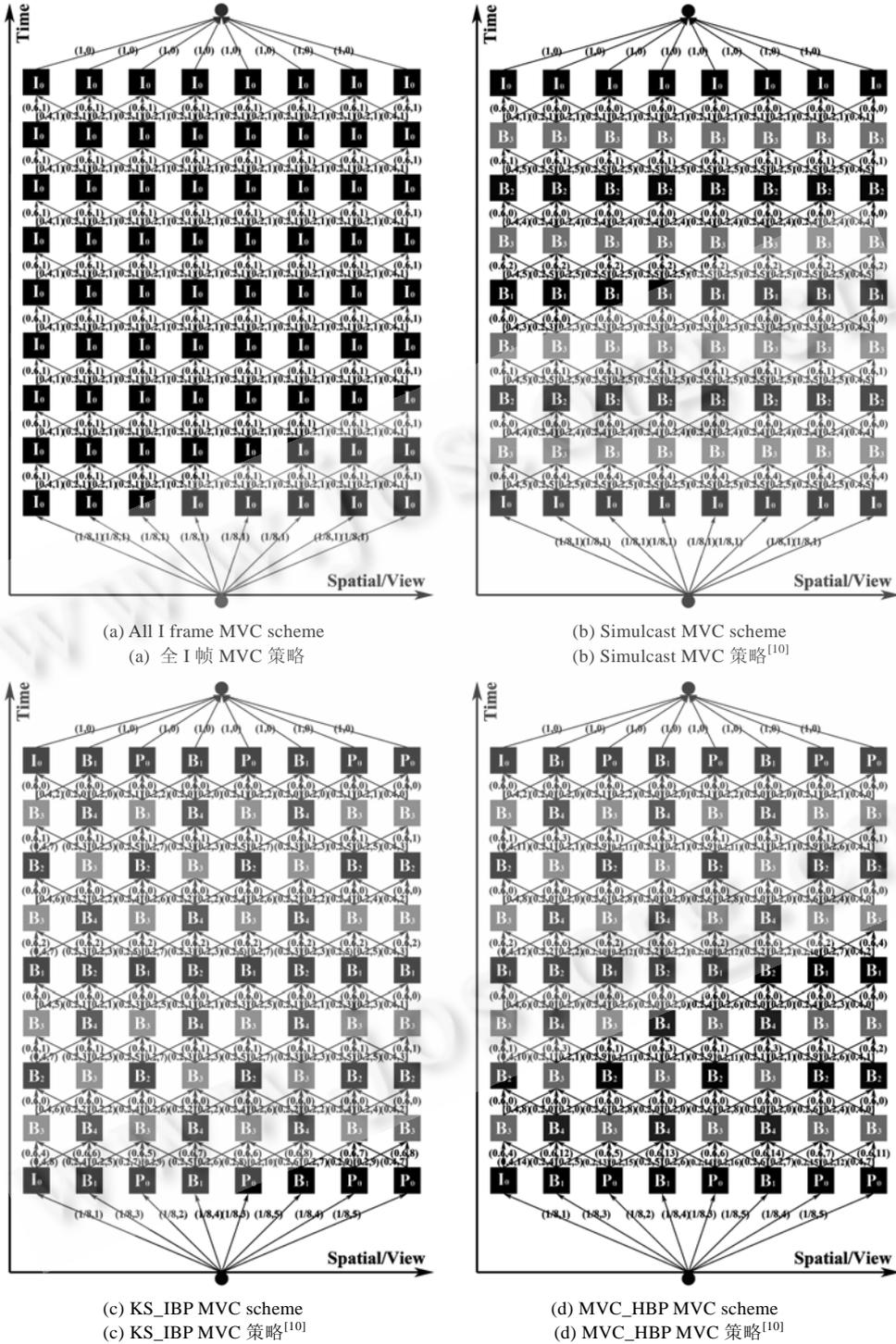


Fig.2 Several MVC schemes

图 2 几种常用的多视点视频编码策略

定理 1. 由有向随机图 G 构成的多视点视频超空间构成了用户以某种交互模式进行观看的所有路径,而用

户进行切换选择的随机事件则构成了从起点到终点的随机路径的概率空间 Ω 这个概率空间一定满足 $p(\Omega)=1$.

证明:以 $N(v_{(i,j)})$ 表示随机图 G 中顶点 $v_{(i,j)}$ 的邻域,其中,下标 j 表示视点坐标, i 表示时间坐标.以 $p_{(i,j),k}$ 表示从当前帧 $v_{(i,j)}$ 切换到下一帧 $v_{(i+1,k)}$ 的概率,其中, $k \in N(v_{(i,j)})$.由动态规划及概率论可以得到下面的递推方程:

$$P_{i,j} = \sum_{k \in N(v_{(i,j)})} P_{i+1,k} P_{(i,j),k},$$

其中, $p_{m,1}=1, j \in [1, n], i \in [0, m]$.通过简单的计算可以得到 $p_{\text{start point}}=1$,即从起点至终点所有的路所构成的概率空间,其概率为 $p(\Omega)=1$. □

从上面的证明过程可知,若能够得到所有的从当前帧切换到下一帧的解码帧数,则可以计算出整个概率空间的解码帧数学期望,即得到这个超空间的随机访问解码帧期望代价.为了便于分析,本文对一个 GOP 进行了分析.由于一个 GOP 也能对应一个超空间,对其分析与对整个多视点视频超空间进行分析是等价的.

对于一个 GOP 而言,其组织结构和预测模式决定了从当前帧切换到下一帧的解码帧数.如图 2 所示,由于 GOP 的组织结构和预测模式不同,从当前帧切换到下一帧的解码帧数各不相同.图中每条边带有两个参数:一个是当前帧切换到下一帧的概率;另一个是从当前帧切换到下一帧的预解码帧数.如果能够计算出所有从起点至终点的随机路径其对应的解码帧代价,则可以对这种 GOP 组织结构的随机访问性能做出一个正确的评价.

随机访问性能评价函数:若以 P 表示 G 中的路,则 P 出现在 G 中的概率为 $P_\varphi = \prod_{e_{ij} \in \varphi} p_{ij}$,而其解码代价为

$W_\varphi = \sum_{e_{ij} \in \varphi} w_{ij}$,其中, w_{ij} 表示边 e_{ij} 上的解码代价.在此基础上,建立 GOP 的随机访问性能评价函数为

$$F_{0,0} = E(\Omega) = \sum_{\varphi \in \mathcal{G}} E(\varphi) = \sum_{\varphi \in \mathcal{G}} P_\varphi W_\varphi = \sum_{\varphi \in \mathcal{G}} \left(\sum_{e_{ij} \in \varphi} w_{ij} \times \prod_{e_{ij} \in \varphi} p_{ij} \right) \quad (3)$$

$F_{0,0}$ 表示从起点至终点的期望解码帧数,即支持用户交互的期望解码帧数.这个期望值反映的是该编码结构支持用户进行随机访问的性能.值得注意的是,建立在概率空间 Ω 基础上的式(3)是一个普适性评价公式,与交互模式和 GOP 结构无关.因此,只要确定了交互模式和 GOP 结构,即意味着式(3)中的 w_{ij} 和 p_{ij} 得到确定,即可使用式(3)进行随机访问性能的评价.以图 2 所示的相邻视点切换的交互模式为例,设定用户在当前帧观看同一视点的下一帧的概率为 0.6,切换到左右相邻视点的概率为 0.2,于是得到图 2 中的带权有向图.图中各边的权重 w 由视频编码策略所决定.通过式(3)可以得到用户以交互式的方法从起点至终点随机访问的性能,见表 1.为了能够方便地在几种规模不同的 GOP 之间进行比较,表中还列出了 $F_{0,0}$ 相对于该 GOP 规模的相对比.显然,由于 $F_{0,0}$ 或者 $F_{0,0}/m_1$ 值表示的是支持随机访问的解码帧代价,所以其值越小,该 GOP 的组织结构以及预测模式对随机访问的性能就越好.类似地,当权重 w 表示服务器端为了支持用户的复杂交互而需要开辟的内存空间时,若以 $F_{m-0,0}$ 表示开辟内存空间期望的大小,则 $F_{m-0,0}$ 也可以通过式(3)进行计算.本文所讨论的 9 种编码策略所对应的 $F_{m-0,0}$ 结果见表 1.

Table 1 Random accessibility of several MVC schemes and SP improved MVC schemes

表 1 几种多视点视频编码策略以及 SP 改进多视点视频编码策略的随机访问性能

GOP structure	$F_{0,0}$	$F_{0,0}/m_1$	$F_{m-0,0}$	$F_{m-0,0}/m_1$	$E(X_m)$	$E(X_j)$
All I frame	9	1	0	0	1	0
Tree mode	7.75	1.94	3.75	0.94	3.5	2.5
Simulcast	17	1.89	12.05	1.34	3.67	2.67
Transposed	23	4.60	18.20	3.64	13	12
Improved transposed	23	4.60	18.58	3.72	13	12
KS_PIP	19.61	2.18	13.22	1.47	7.22	6.22
KS_IPP	22.95	2.55	16.62	1.85	8.75	7.75
KS_IBP	22.1	2.46	15.7	1.74	6.53	5.53
MVC_HBP	27.79	3.09	21.99	2.44	9.31	8.31
SP transposed structure (1)	15	3.00	10.77	2.15	5	4
SP transposed structure (2)	13.93	2.79	9.86	1.97	5.64	4.64
SP improved transposed structure	13.42	2.69	9.70	1.94	6.28	5.28

表 1 分别用 $F_{0,0}, F_{m-0,0}, E(X_j)$ 和 $E(X_m)$ 表示 GOP 随机访问性能、GOP 支持随机访问内存大小、随机访问单

帧性能和随机访问单帧内存消耗大小,用 $F_{0,0}/m_1$ 和 $F_{m-0,0}/m_1$ 表示相对 GOP 随机访问性能和相对 GOP 支持随机访问内存大小.由表 1 可见,全 I 帧 GOP 结构对随机访问评价的各项参数表现均为最优的,串联 GOP 结构则是相对最差的.然而,采用全 I 帧 GOP 结构对多视点视频进行编码是不现实的,其中最主要的问题在于该编码方式受到通信带宽的限制.同时,Simulcast 编码结构^[10]与 MPEG 早期对立体视频进行编码的方案相似,但是,由于这种编码方案没有考虑帧间的视差关系而导致压缩比率不高.考虑到编码效率的需求,则必须增加视点间的预测依赖关系.MVC_HBP 结构虽然随机访问性能比较差,但其编码性能是最优的.KS_IPP,KS_IBP,KS_PIP^[10]的随机访问性能较 MVC_HBP 要好,但其编码性能却又相对较差.因此,各编码结构在不同的条件下具有各自的优缺点.如何在实际应用中选择合适的多视点编码方案,平衡实际当中的各种需求,选择出具有优良性能的编码方案使之得到应用等各种问题,就成为了人们需要进一步考虑的对象.

3 多视点视频编码策略评价模型

由图 2 以及表 1 的计算结果可以发现:(1) 当用户交互模式以及访问概率确定后,决定随机访问性能优劣的参数只有 w ,说明编码策略的帧间依赖关系会对随机访问性能产生直接的影响;(2) 决定 w 大小的是多视点视频 GOP 的预测结构,即编码过程中所采用的预测模式和编码策略.一般情况下, w 的大小与对应的预测模式所产生的比特率成反比关系.

现有的多视点系统要求编码端在保证视频质量的基础上实现高压缩比,即得到好的率失真性能;而在解码端,要求用户在使用过程中能够实现低延迟的交互式观看.这两种要求本质上是一对矛盾体,因为高压缩比往往意味着帧间依赖关系密切,从而使权变量 w 增大,进而导致随机访问性能低下;随机访问性能高的方案往往在带宽上受到很大的限制.如何在这一对矛盾中找到平衡是一个值得研究的问题.预测模式的选择对随机访问性能的优劣起着至关重要的作用,而预测模式的选择又与实际应用背景要求的通信带宽直接相关.同时,也需要考虑客户端的随机访问性能以及服务器端支持随机访问所付出的代价.根据这些相互制约的条件,得到用于分析多视点视频编码策略优劣的评价模型.

多视点视频编码策略评价模型:设 $E(X_m), E(X_f), F_{0,0}, F_{m-0,0}$ 分别表示某编码策略对应的单帧随机访问参数和复杂随机访问性能参数,设 B, F, F_m, E_f, E_m 分别表示实际应用中的带宽、复杂随机访问以及单帧随机访问的限制, $B(\bar{w})$ 表示指定编码策略所产生的码率, s_i^-, s_i^+ 分别为对应于各个随机访问性能参数,对应于性能限制约束条件的正、负偏差变量,且有 $s_i^+ \times s_i^- = 0$.根据多目标非线性数学模型的基本建模过程可知,若希望编码策略对应的单帧随机访问参数、复杂随机访问性能参数和码率都能满足事先规定的所有目标,则多视点视频编码策略的评价模型可以表达为

$$\begin{aligned} \min z &= s_1^+ + s_2^+ + s_3^+ + s_4^+ + s_5^+ \\ \text{s.t.} &\begin{cases} B(\bar{w}) + s_1^- - s_1^+ = B \\ F_{0,0} + s_2^- - s_2^+ = F \\ F_{m-0,0} + s_3^- - s_3^+ = F_m \\ E(X_f) + s_4^- - s_4^+ = E_f \\ E(X_m) + s_5^- - s_5^+ = E_m \end{cases} \\ &(\bar{w} \in N) \geq 1, s_i^- \geq 0, s_i^+ \geq 0 \end{aligned} \tag{4}$$

目标方程仅由偏差变量组成,则说明该模型如果最优解存在,则最优值一定为 0.由于有偏差变量的限制,所以该模型的意义为:只要编码策略能够满足所有的约束条件,则该策略即为本模型的最优解.显然,该模型的最优解若存在,则必定不唯一.由于在交互模式确定之后,即用户进行切换的概率确定之后,编码策略是否为最优解仅与权重 \bar{w} 有关,因此,该模型中需要深入讨论的变量仅为权重 \bar{w} .

上述多视点视频编码策略评价模型除了对视频编码策略进行评价之外,还可以作为多视点视频编码端进行预测模型策略选择的一个工具.如全 I 帧 GOP 编码结构虽然具有最小的随机访问代价,但是由于其在绝大多数情况下不满足评价模型(4)的约束方程 $B(\bar{w}) \leq B$,所以不可能被选为编码模型进行使用.

4 多视点视频编码策略评价模型的应用

为了提高多视点视频的随机访问性能,比较常用的做法是在编码的过程中对适当位置的帧进行 SP(switch prediction)帧编码,以方便用户在这个位置进行视点切换^[11].如何确定在加入 SP 帧的同时不突破实际带宽的限制并达到较好的随机访问性能,又成为人们的一个研究课题.在这里,采用多视点视频编码策略评价模型可以对编码策略中 SP 帧的摆放位置进行设计,在实验之前就对该模型是否值得继续研究做出一个先验性的判断.

从表 1 可以看出,串行结构和改进的串行结构虽然具有易于编码实现的优点,但是其随机访问性能很差.对其进行 SP 帧改造,如图 3 所示.改造后的各个评价结果见表 1.可以发现,改造后的结果对随机访问性能有了极大的提高.下面只需对其进行码率控制和优化设计,即可在一些带宽限制不高的环境中加以采用.

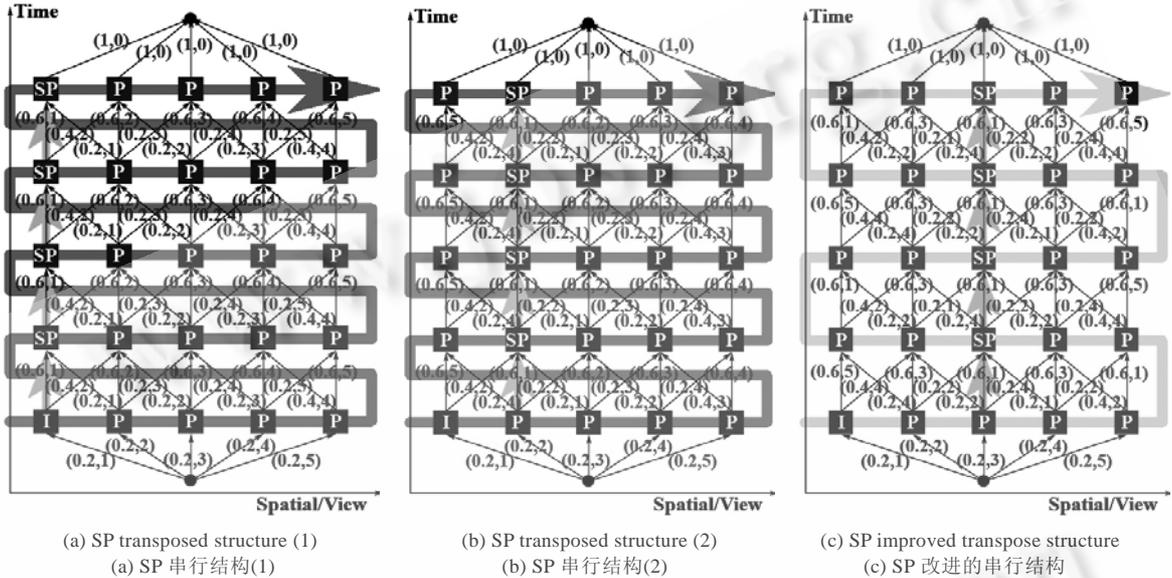


Fig.3 SP frame design for transposed and improved transposed MVC structures

图 3 对串行结构和改进的串行结构进行 SP 帧设计

5 结论

多视点视频编码是近些年来视频处理领域中的一个研究热点.多视点视频编码在追求高率失真效果的同时,还需要同时兼顾到解码端的随机访问、与用户的交互等性能.本文首先定义了 3 种复杂的用户与服务器交互模式,利用随机图论、超空间理论、动态规划以及概率分析的相关理论知识,得到了对解码端支持用户随机访问性能进行评价的随机访问评价函数.用此函数可以对不同的 GOP 组织结构进行无类别的评价,所得结果可以在相对性能的条件下进行比较.利用解码帧的期望数进行随机访问性能的评价,当出现不同的运行环境和条件时,它们之间的差异只会相差常数倍.在本文的实验阶段选取了 9 种比较特殊和处于研究热点的 MVC 编码方案进行了讨论.同时,利用所提出的评价函数再加上相关的约束条件,就得到非线性多目标的多视点视频编码策略评价模型.利用这个评价模型,可以得到客观评价指标,并对不同的预测模型进行筛选,或进行先验性的优化设计.实验与设计结果表明,在事先不进行编码的情况下,即可针对不同的实际应用情况对 MVC 编码方案进行优化和筛选,达到事半功倍的效果.

References:

- [1] ISO/IEC JTC1/SC29/WG11. Applications and requirements for 3DAV. Doc. N5877, Trondheim, 2003.
- [2] ISO/IEC JTC1/SC29/WG11. Requirements on MVC v.4. MPEG 2005/N7282, 2005.

[3] Ramanathan P, Girod B. Random access for compressed light fields using multiple representations. In: Proc. of the IEEE 6th Workshop on Multimedia Signal Processing (MMSp 2004). Washington: IEEE Computer Society, 2004, 383–386.

[4] Ramanathan P, Girod B. Rate-Distortion optimized streaming of compressed light fields with multiple representations. In: Proc. of the Int'l Packet Video Workshop. 2004. http://msw3.stanford.edu/~pramanat/Papers/tmm_lfstreaming.pdf

[5] Kimata H, Kitahara M, Kamikura K, Yashima Y. System design of free viewpoint video communication. In: Proc. of the 4th Int'l Conf. on Computer and Information Technology (CIT 2004). Washington: IEEE Computer Society, 2004. 52–59.

[6] Lou J, Cai H, Li J. A real-time interactive multi-view video system. In: Proc. of the 13th ACM Int'l Conf. on Multimedia (ACM MULTIMEDIA 2005). New York: Association for Computing Machinery, Inc., 2005. 161–170.

[7] Liu Y, Huang Q, Ji X, Zhao D, Gao W. Multi-View video coding with flexible view-temporal prediction structure for fast random access. In: Advances in Multimedia Information Processing-PCM 2006. LNCS 4261, 2006. 564–571.

[8] Bollobas B. Modern Graph Theory. New York: Springer-Verlag, 1998.

[9] Yang Y, Jiang G, Yu M, Li F, Kim Y. Hyper-Space based multiview video coding scheme for free viewpoint television. In: Proc. of the Picture Coding Symp. (PCS 2006). Beijing: Elsevier Engineering Information Inc., 2006.2–23.

[10] Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG. Comparative Study of MVC Prediction Structures. 2007.

[11] Guo X, Lu Y, Gao W, Huang Q. Viewpoint switching in multiview video streaming. In: Proc. of the Int'l Symp. on Circuits and Systems. 2005. 3471–3474.



杨铀(1978—),男,广东广州人,博士生,主要研究领域为视频编码,信号处理,图论与组合优化.



郁梅(1968—),女,博士,教授,主要研究领域为多媒体信号处理,编码与通信,计算机视觉及应用.



蒋刚毅(1964—),男,博士,教授,博士生导师,CCF高级会员,主要研究领域为图像处理与视频压缩,多媒体信号处理与通信,基于视觉的智能控制.



曹德侠(1977—),女,讲师,主要研究领域为数学分析,数学模型.